UNIVERSITY OF ENGINEERING AND TECHNOLOGY PESHAWAR

Department of Computer Science and Information Technology

(DATA SCIENCE)

Mid Term Examination Fall-2022

Paper: BSI-231 Statistics and Probability

Time: 2 Hrs

Note: Attempt ALL questions.

Max. Marks: 100

21PWDSC0029

1. (a) **List** each of these as Nominal, Ordinal, Interval or Ratio level of measurement and why. (20+5)

   i. A course is being rated by students on a scale of 1 star to 5 stars. Interval
   ii. A physiotherapist notes the pain level on a scale of 1 to 10. Interval
   iii. A teacher notes down the weight of each student in the class. Ordinal
   iv. One column in a data table is the daily temperature in degrees Celsius. Ratio
   v. One column in a data table is the gender of the employee (e.g. Male, Female, Undefined). Nominal
   vi. One column in a data table is the employee's designation (e.g. CEO, Manager, Supervisor) Nominal
   vii. Highway route numbers. Nominal
   viii. Political party voted for in the last election (e.g. party X, party Y, party Z) Nominal
   ix. Number of employees at a company. Nominal
   x. Income categorized as ranges ($30-39k, $40-49k, $50-59k, and so on). Interval
   Note: 1 mark for name and 1 mark for reason.

   (b) (i) A survey will be given to 100 students randomly selected from the freshmen class at Lincoln High School. **Name** is the population. Random
   (ii) A survey will be given to 100 students randomly selected from the freshmen class at Lincoln High School. **Name** is the Sample. Random
   (iii) Every seventh customer entering a shopping mall is asked to select her or his favourite store. **Name** the sampling type. Cluster
   (iv) In a large school district, all teachers from two buildings are interviewed to determine whether they believe the students have less homework to do now than in previous years. **Name** the sampling type. Stratified
   (v) Selection of 10 different sites by Civil Engineering students for doing surveys. **Name** the sampling type. Systematic

2. The United States has an aging infrastructure as witnessed by several recent disasters, including the I-35 bridge failure in Minnesota. Most states inspect their bridges regularly and report their condition (on a scale from 1-17) to the public. Here are the condition numbers from a sample of 30 bridges in New York State: (25)

   5.08 5.44 6.66 5.07 6.80 5.43 4.83 4.00 4.41 4.38
   7.00 5.72 4.53 6.43 3.97 4.19 6.26 6.72 5.26 5.48
   4.95 6.33 4.93 5.61 4.66 7.00 5.57 3.42 5.18 4.54

   Find the sample Mean, Median and sample Standard Deviation of these condition numbers.

$$\sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}}$$

$$\frac{S.D}{Mean} \times 100 \% \quad \text{co-efficient}$$

Cluster
Random Stratified
Systematic

3. Over a period of 60 days the percentage relative humidity in a vegetable storage building was measured. Mean daily values were recorded as shown below: (25)

6,7,8,9

885 − 721

```
60 63 64 71 67 73 79 80 83 81
86 90 96 98 98 99 89 80 77 78
71 79 74 84 85 82 90 78 79 79
78 80 82 83 86 81 80 76 66 74
81 86 84 72 79 72 84 79 76 79
74 66 84 78 91 81 64 76 78 82
```

(a) **Derive** a stem-and-leaf display with at least five stems for these data.
(b) **Find** First, Third Quartiles and 4th decile.

$\dfrac{164}{1606}$

$\dfrac{1134}{20}$ *56.7.*

4. Water quality measurements are taken daily on the River Ouse at Clapham, England. The concentrations of chlorides and phosphates in solution, given below in milligrams per litre, are determined over a 20-day period.

86.9 (25)

Chloride: 64.0, 66.0, 64.0, 62.0, 65.0, 64.0, 64.0, 65.0, 65.0, 67.0, 67.0, 74.0, 69.0, *1.8.45*
68.0, 68.0, 69.0, 63.0, 68.0, 66.0, 66.0

Phosphate: 1.31, 1.39, 1.59, 1.68, 1.89, 1.98, 1.97, 1.99, 1.98, 2.15, 2.12, 1.90 1.92, 2.00, 1.90, 1.74, 1.81, 1.86, 1.86, 1.65

Choose which one has the greater variation among the Chloride and Phosphate.

60
63
64 =
66
67
71
72
73
76
77
78 =
79 —

= 3

$\dfrac{Q_2 - Q_1}{Q_2 + Q_1}$

$\dfrac{j(n+1)}{4.}$

DEPARTMENT OF COMPUTER SCIENCE & IT, UET PESHAWAR

PAPER TITLE:

COMPUTER ORGANIZATION AND ASSEMBLY LANGUAGE

MID TERM EXAM BS (CS) & DS 3RD SEMESTER (FALL-2022)

TIME: 2 HOURS   TOTAL MARKS: 100

ATTEMPT ALL QUESTIONS   INSTRUCTOR: ALAUDDIN

Q # 1: What is the difference between "Compiler" and "Assembler".   [4]

....

....

....

Q # 2: Enlist four basic functionalities and four classic components of a computer.   [4+4]

....

....

....

....

[4+4]

Q # 3:   Convert :

   a. 45_{10} to hex.

   b. 39_{10} to Binary.

Q # 4:   To address 20K bytes memory how many address lines are required? Calculate.   [2]

....

Q # 5:   Two address buses has a width of 10 and 14 lines. What will be the address of the last location if   [4+4]

it starts from 0000.Calculate.

....

Q # 6 : Write down MIPS code for the following statements.

f = (g + h) – (i + j). If the variables f, g, h, i, and j are assigned to the registers $s0, $s1, $s2.$s3, and $s4. respectively.

f → s0
g → s1
h → s2
i → s3
j → s4

add $t0, $s1,$s1[10]
add $t1 ,$s3,$s4.
sub $s0 ,$t0,$t1.

$26
26

Q # 7 : Write MIPS code to transfer value "100" from the location shown in figure to register t0 (base address in $s3). [4]

100	
10	
101	
1	

Byte Address   Data

Q # 8: Below given figure represents little-endian or big-endian? Explain why. [2]

Byte #

3	2	1	0

Q# 9: Write down MIPS code for A[8]=h + A[2]; if h is in $s2, base address of A is in $s3. All the values should be stored in $t0. [12]

$s2

Base address
$s3.

$t0

lw

lw $s3 , 2.
add $t0 , $s2,$s3
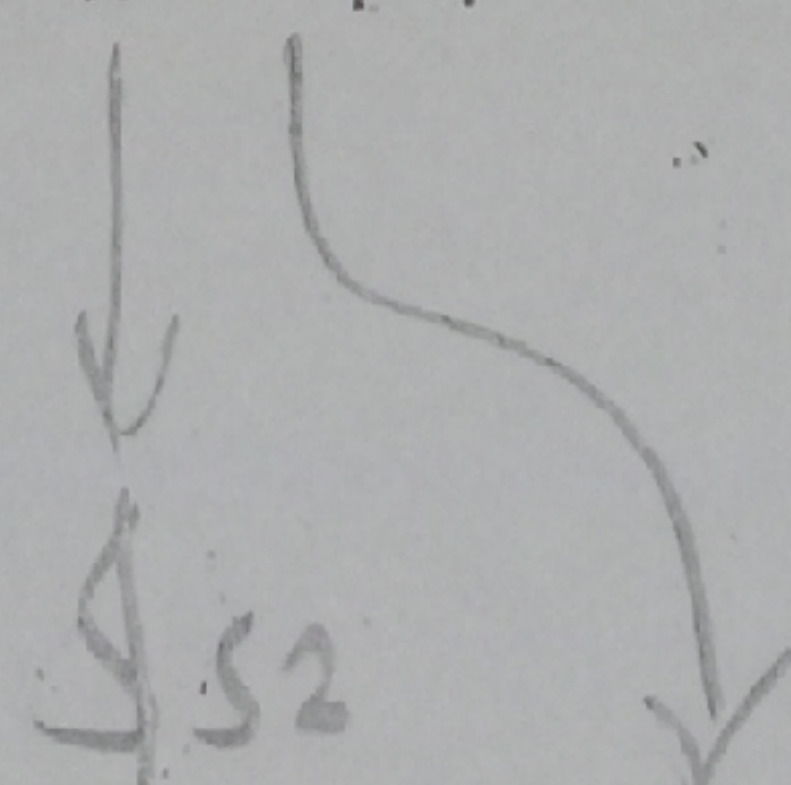
Q# 10: Negate 12.

[4]

10

@ `101` →

$\frac{1}{10}$

Q # 11: For below given two machine codes (each of 32 bits), write down correct MIPS instructions.   [12]

0	16	19	17	0	34	and
8	16	17		100		100.

↓   ↓   ↓   ↓

addi   S2 . S1

Q # 12: Write MIPS code that stores 5 values in memory and compare all of them with k==20. If any value matches k== 20 it will multiply that value with 4 and will come out of the loop. If no value matches, the last value compared will be incremented by 20. Use $S0 for k.$S1 for loop counter and so on..(Do not use ant temporary registers).   [14]

0| 0|

Q # 13: Write a function (Procedure) using MIPS that converts a lower case character to uppercase and Uppercase to lowercase . Any other value is returned unchanged.

[12]

$s1 9 $s'

$l'

btts

blt 9 $s1

bgt $s2

jr $s1

π jren

26
26
2

21

STUDENT REG #: _____

PAPER TITLE: Data Structure & Algorithm
MID TERM EXAM
TIME: 2 HOURS
ATTEMPT ALL QUESTIONS

BS (CS) – 3rd SEMESTER
TOTAL MARKS: 100
INSTRUCTOR: SADIQ UR REHMAN

Question 1. Answer the following question   (20 MARKS)

i)   Why we need Data Structure? What are the key properties of an algorithm?

ii)   What is Abstract Data Type (ADT)? Why we need ADT

iii)   Write C++ Code for in Insertion in Array.

iv)   Write C++ Code for Binary Search using link list.

v)   Write C++ Insertion Sort.

Question 2.

Part (a) Differentiate between Merge sort and Quick Sort. Also write C++ Code for Quick Sort.
                                                              (10 MARKS)

Part (b) How Quick Sort works. Demonstrate your answer with help of the following example
                                                              (10 MARKS)

Let's consider an array with values {9 , 7, 5, 11, 12, 2, 14, 3, 10, 6}
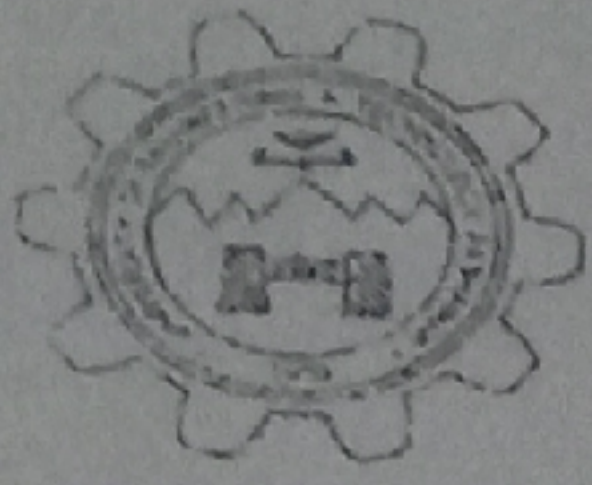
Question 3.

Part (a) Explain Comparison based sorting algorithms. Can we do better time complexity of Comparison
based sorting algorithms than O(n log n).   (10 MARKS)

Part (b) What is Counting Sorting. Write algorithm for Counting sort. Demonstrate the algorithm with

help of the following example

Index	1	2	3	4	5	6	7	8	9	10	11	12
Data	3	8	1	3	9	4	5	3	9	7	6	2

STUDENT REG #: _____

Question 4.

Part (a) Explain Node of a link list. Write code in C++ for a node. List the advantages of doubly link list over single link list.

(10 MARKS)

Part (b) Write the code in C++ in which user enter a value and program will find the value from the singular link list.

(10 MARKS)

Question 5.

Part (a) What do you know about stack? Describe the purpose of each of the following stack operations, also write C++ code for each operation.

(10 MARKS)

i)	Push ()	ii)	Pop ()	iii)	Peek ()
iv)	Isfull ()	v)	Isempty ()		

Part (b) Write the algorithm to convert infix expression to postfix expression. Convert the following infix expression into its postfix form using stack operations

(10 MARKS)

i) $a * (b * c) + (d * e)$

ii) $(70 + 20) * (45 - 17) / (74 + 17)$

****

21PWDSC0029

COURSE TITLE: TECHNICAL REPORT WRITING & COMMUNICATION SKILLS
3rd SEMESTER BS (CS)
TIME ALLOWED: 2 HOURS

Attempt all questions:

Q.1. Read the following introduction paragraph carefully and answer the questions bellow: (10)

Media, particularly television, has a significant role in the everyday lives of children. The way media influences children can be both positive and negative. Quality television and written texts can provide positive role models and opportunities for learning. However, media showing excessive or gratuitous behaviors can be harmful to the development of children's attitudes and values. Initially, a discussion about the many types of media will caution that it cannot be assumed media has the same effects on all children. Further, it will be shown that media can be actively used to achieve beneficial outcomes among children, including the range of educational television tailored for children that build social skills and introduce children to diverse themes and topics. In contrast, the negative impacts of media, including exposure to inappropriate material, creating a scene of entitlement, increasing peer pressure and restricting contact children have with real people will be examined. Finally, the amount and type of media, variety and quality of content will be discussed to highlight that media becomes harmful to children when exposure to media is not monitored or when children miss out on other activities; such as exercise and social media.

   i.   Search/think a suitable 'essay title' for the above introduction paragraph.

   ii.   Write a complete 5-paragraph essay in line with the given paragraph.

   iii.   Identify the main points/arguments in the given para and reproduce them.

   iv.   Identify the signal words/transitions in the given para and reproduce them.

Q.2.   a. Name and describe the different components of a rhetoric situation.

     b. Describe the process of critical thinking diagrammatically. (05)

Q.3.   a. Name the different parts of a technical report.

     b. List the four language areas for writing the result section of an article/paper. (05)

Q.4.   a. List the information you have to present in an introduction to an article/paper.

     b. Name the three language areas which are important for writing the methodology. (05)

****

STUDENT REG #: _____

PAPER TITLE: CS-318 INTRODUCTION TO DATA SCIENCE	
MID TERM EXAM	BS (DS) – 3RD SEMESTER
TIME: 2 HOURS	TOTAL MARKS: 30
ATTEMPT ALL QUESTIONS.	INSTRUCTOR: DR. SUHAIL YOUSAF

A credit card is a thin rectangular piece of plastic or metal issued by a bank or financial services company that allows cardholders to borrow funds with which to pay for goods and services with merchants that accept cards for payment. Credit cards impose the condition that cardholders pay back the borrowed money, plus any applicable interest, as well as any additional agreed-upon charges, either in full by the billing date or over time.

A vast majority of businesses let the customer make purchases with credit cards, which remain one of today's most popular payment methodologies for buying consumer goods and services. As of June 2018, there were 7.753 billion credit cards in the world. In 2020, there were 1.09 billion credit cards in circulation in the U.S. and 72.5% of adults (187.3 million) had at least one credit card in this country. In order for the credit card scheme to sustain it is of utmost importance that client should pay back the amount consumed via the credit card timely.

This poses a challenge for the credit card issuing authorities. The challenge is to assess, before issuing the card, the client's financial soundness for paying the debt as per the specified timeline. However, such an assessment is not a straight forward process and involves dealing with uncertainties. So, considering the huge number of credit card users, the main problem for banks regarding credit cards is to sustain the system by issuing cards to suitable clients. Certain clients will not be feasible as they do not have the economic strength to back up this service.

Being a freshly hired data scientist at DatumWoods Solutions: a data science company, you are expected to identify the problem and provide a data science based solution to the problem.

Following the data science methodology, answer these questions (concisely!!!):

1. Identify and phrase the Data Science problem in the given scenario.
2. Discuss the nature of the problem statement described in the previous step as either classification or regression problem and justify your answer.
3. Reflect on the data requirement. What data is needed to answer the question?
4. Reflect on data collection. Where is the data coming from or how to get it?
5. Considering the given scenario, what would you consider to ensure that the data collected is representative of the problem being solved?
6. What additional work would you do to manipulate and work with the data in the context of the given scenario?

The main problem
which is

COURSE TITLE: TECHNICAL REPORT WRITING & COMMUNICATION SKILLS

3rd SEMESTER BS (CS)

TIME ALLOWED: 2 HOURS　　　　　21PWDsc0029　　　　**MARKS: 50**

Q # 1.　a.　　What type of information should be in the Introduction of an article? Discuss　　(05)

　　　　b.　　Name the different language areas which are important in the introduction. Also describe the importance of any two of them.　　(05)

Q # 2.　a.　　What type of information should be in the Abstract of an article? Discuss　　(06)

　　　　b.　　State the reasons for the two models of the Abstract.　　(04)

Q # 3.　a.　　Describe modal verbs. Also state the two problems associated with these verbs.　　(04)

　　　　b.　　Explain the Traditional Outline Format of a technical report.　　(06)

Q # 4.　a.　　Write down the tips for delivering a good presentation.　　(05)

　　　　b.　　Briefly explain the four methods of presenting a speech.　　(05)

Q # 5.　　　Differentiate /describe the following:　　(04+04+02)

　　　i.　　　Proposal and Progress report.

　　　ii.　　Persuasive and Synthesis essay.

　　　iii.　　Thesis statement and Title of technical report.

****

STUDENT REG #: __21PWDSC0029__

PAPER TITLE: Data Structure & Algorithm

FINAL TERM EXAM

TIME: 2 HOURS

ATTEMPT ALL QUESTIONS

BS (DS) – 3rd SEMESTER

TOTAL MARKS: 50 (Weight 50 %)

INSTRUCTOR: SADIQ UR REHMAN

Question 1. What is binary search tree. How it is different from the binary tree? Write an algorithm for pre-order, in-order and post order traversal of a tree. Construct binary tree for the following data   (14 , 2 , 11 , 1 , 3 , 10 , 30 , 7 , 40)   [10 Marks]

Question 2.

Part(a) What are the probing techniques for collision in closed hashing. Explain in detail the quadratic probing technique with example. Using the following data N = 100; U = [1; 2; : : : ; 100]; m = 7; $h(x) = (x \bmod 7)$; Ś = [3; 19; 4; 22; 17].   [10 MARKS]
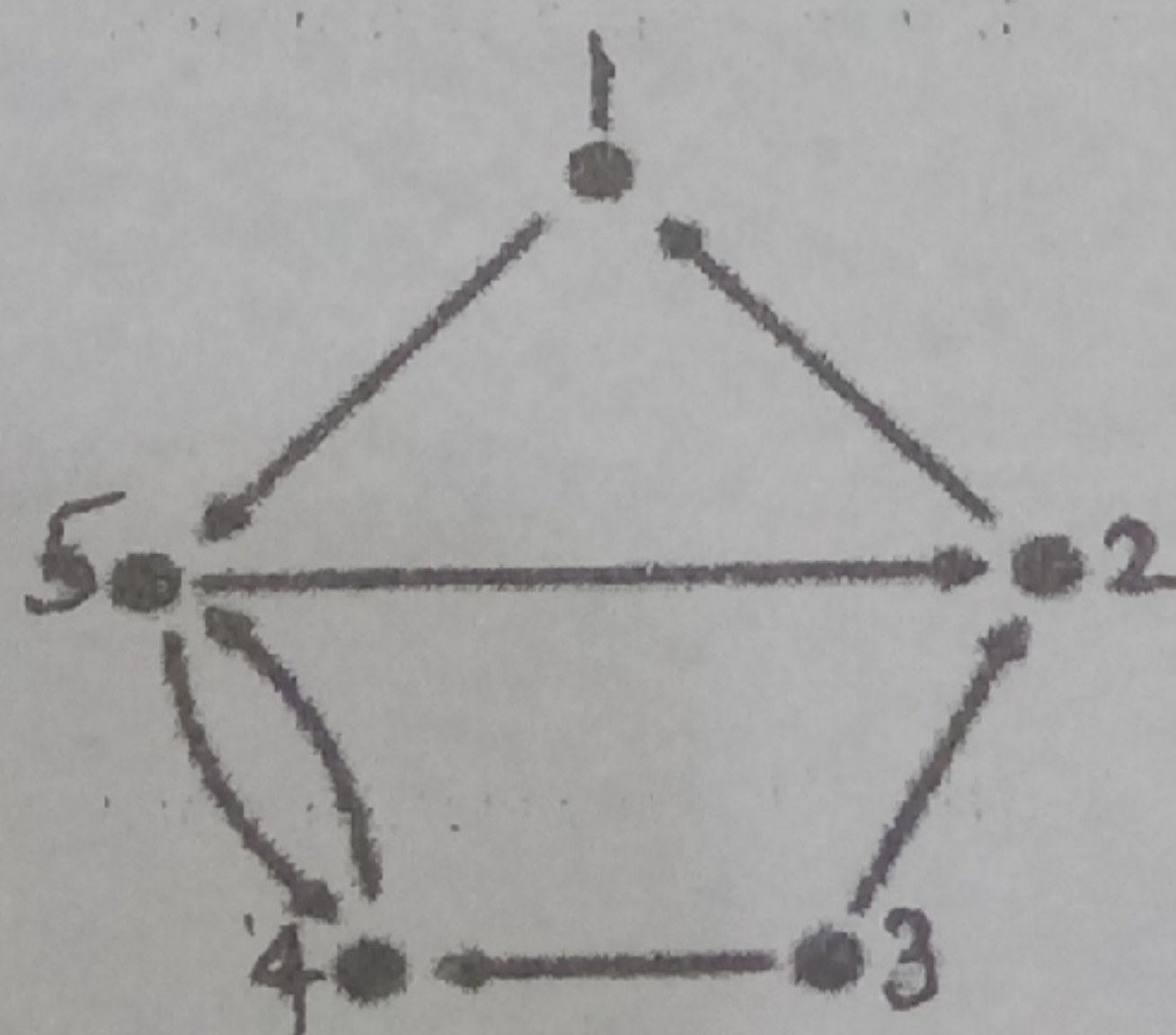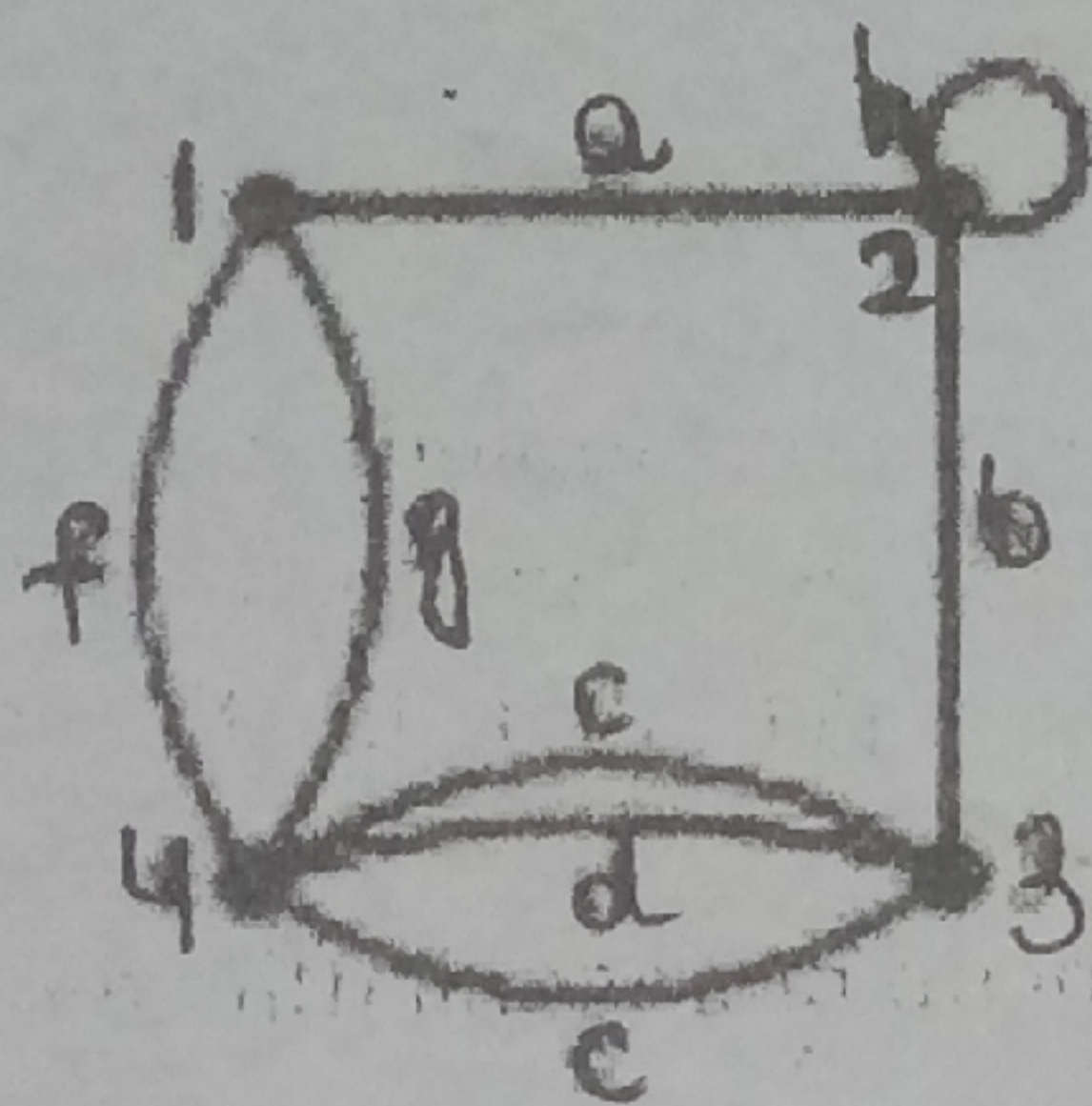
Part (b) What is the problem with linear and quadratic probing in hashing? What is open hashing.

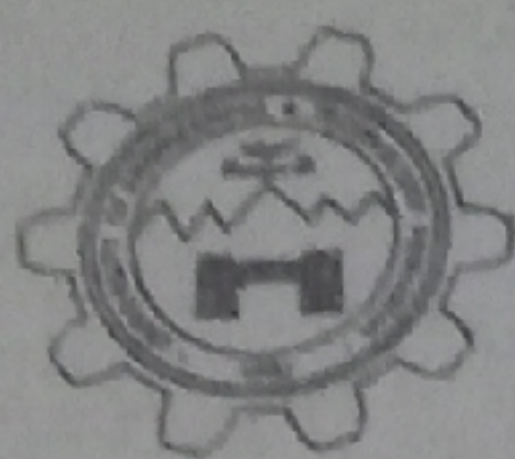Explain the separate chaining technique with example.

[10 MARKS]

Question 3.

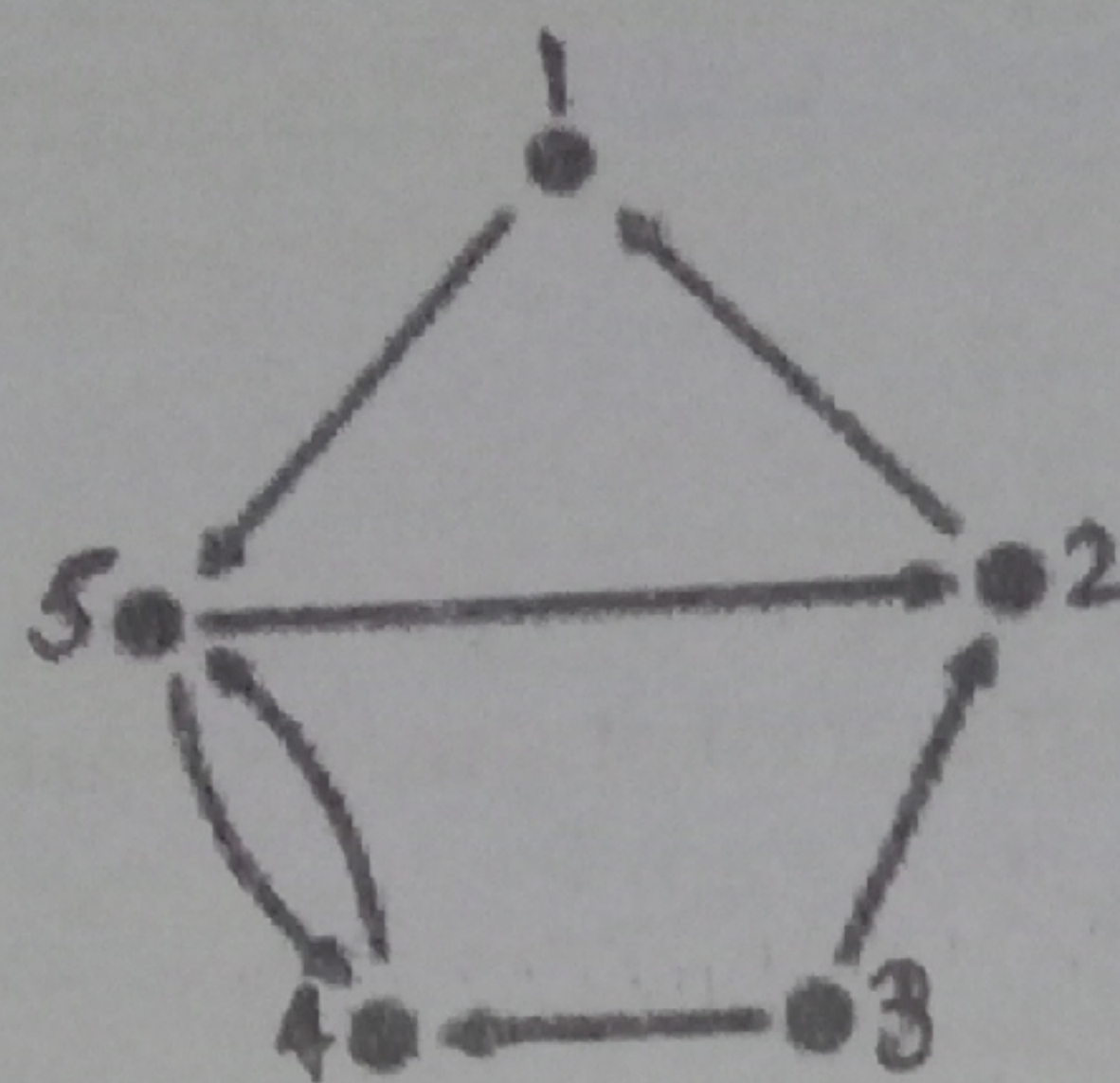Part(a) Write down the adjacency matrix for the following graphs.   [10 Marks]

i.

ii.

a b c d e b g

[05 Marks]

Part(b) Find out the in degree, out degree and degree of all the vertex.

Part (c) Find out the total Number of edges in K_n where n = 5 in a complete graph.   [05 MARKS]

$$\frac{n!}{r!(n-r)!}$$

PAPER TITLE: COMPUTER ORGANIZATION AND ASSEMBLY LANGUAGE

MID TERM EXAM BS (CS):   3RD SEMESTER (FALL-2022)

TIME: 2 HOURS   TOTAL MARKS: 100

ATTEMPT ALL QUESTIONS   INSTRUCTOR: ALAUDDIN

Q # 1: Write down a procedure in MIPS assembly for the following C code. Also use stack. The parameter variables g, h, i, and j for the procedure correspond to the argument registers $a0, $a1, $a2, and $a3, and f corresponds to $s0. Push all the parameters in the stack and then pop them at the end of the procedure. The procedure must also return a value to the caller. Show the stack before, during, and after the procedure call.   [15]

```
int leaf_example (int g, int h, int i, int j)
{
    int f;
    f = (g + h) - (i + j);
    return f;
}
```

Q # 2: Fill the blank boxes with the code for each instruction, all the values must be in hexa decimal format. [20]

```
Loop: sll  $t1, $s3, 2  => 80000
      add  $t1, $t1, $s6 =>
      lw   $t0, 0($t1)   =>
      bne  $t0, $s5, Exit =>
      addi $s3, $s3, 1   =>
      j    Loop           =>
Exit: ...
```


Q # 3: Draw a figure that illustrates the functionality of Program Counter.   [6]

Q # 4: Show the datapath for the "Store" instruction. Clearly showing the data path flow, components used , the control signals and the buses width.   [14]

Q # 5: Write down the values of zero flag and PCSrc for the following instructions.   [10]

Op	Zero flag	PCSrc
R-type		
J		
BEQ		
BEQ		
BNE		
BNE		
Other than Jump or Branch		

Q # 6: Show how a forwarding mechanism is implemented to solve the Read After Write – RAW Hazard for the following instructions along with the necessary control signals. Also draw a graph (Time(x –axis) versus Program execution cycle(y-axis)) to represent the forwarding mechanism. [15]

sub $s2, $t1, $t3

add $s4, $s2, $t5

or $s6, $t3, $s2

and $s7, $s6, $s2

sw $t8, 10($s2)

Q # 7: A computer system has a 64 MB main memory. It also has 8K-Byte cache organized as a 2-way set associative, and 64 bytes per block. Calculate the number of bits in the tag, Set index and Byte Offset fields of the memory address format. [14]

Q # 8: What is the difference between a virtual machine and a container. [6]

665536

STUDENT REG #: 21PWDSC0029

PAPER TITLE: CS-318 INTRODUCTION TO DATA SCIENCE

FINAL TERM EXAM

TIME: 2 HOURS

ATTEMPT ALL QUESTIONS.

BS (DS) – 3RD SEMESTER

TOTAL MARKS: 100

INSTRUCTOR: DR. SUHAIL YOUSAF

1. What is the difference between analysis and synthesis?
2. What is the purpose of data analysis step in Data Science?
3. What is meant by a model?
4. Where is model in $y=f(x)$?
5. *What do we determine through ML in $y = mx + c$* —> cofficient
6. What is the purpose of data wrangling?
7. How do you deal with missing values in data?
8. What is the role of binning in data analysis?
9. What is the purpose of exploratory data analysis? —> Correlation (statistics)
10. Explain the role of pivot in EDA. How Heat map and pivot are related?
11. Explain the premise: Correlation does not imply causation.
12. What is meant by Regression? Write down various variants of Regression.   Cor
13. Explain cross-validation in model evaluation.
14. What is the different between under-fitting and over-fitting?
15. Name different types of Machine Learning.
16. Write down names of Python libraries for scientific computing, visualization, and algorithms.
17. What is data normalization? Why it is important? Write down three basic normalization approaches.
18. Explain the structure of Box Plot through a labeled diagram? How outliers are treated by Box Plot?
19. What is the purpose of Pearson correlation? Comment on the meaning of Correlation coefficient and p-value.
20. How categorical values are converted into quantitative data? Why this conversion is needed?
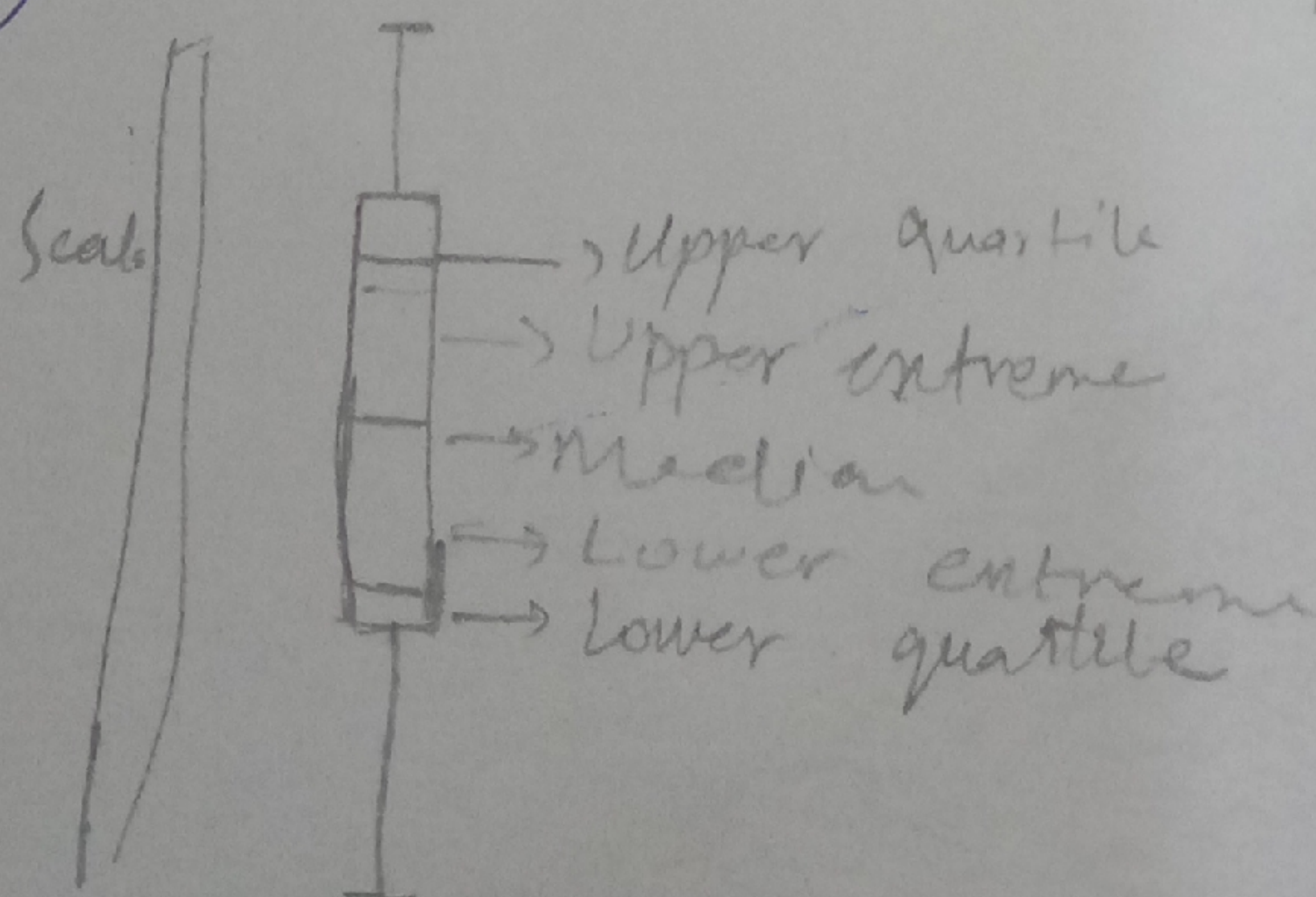
Time Allowed: 3 hours   2/PWDSC 0029.   Max Marks: 100

Note: Attempt ALL questions: Calculator Exchange is not allowed.

Q1.   a) A winner and a runner-up are decided in a tournament of four players, one of whom is Terica. If all the outcomes are equally likely, what is the probability that (i) Terica is the winner? (ii) Terica is either the winner or the runner-up?

b) The probability that A will be alive after 10 years to come is 5/7 and for B it is 7/9. Find the probabilities that: (i) both will be alive (ii) only A will be alive (iii) both will die.   (10+10)

(handwritten margin notes:)
$$\frac{n!}{r!(n-r)!}$$

$$\frac{4!}{2!(4-1)!}$$

$$4 \times 3!$$

$$\pm 1 \cdot (3)!$$

$2/7$ $P(A \cup B) = P(A) +$

$P(B).$

Q2.   a) A random variable X has a pdf given by

(handwritten:) $5/14$ $7/18$ $10/63$ $= 4/9$

$$f(x) = \begin{cases} \dfrac{15}{64} + \dfrac{x}{64}, & -2 \le x \le 0 \\[2mm] \dfrac{3}{8} + ax, & 0 < x \le 3 \\[2mm] 0, & elsewhere \end{cases}$$

(i) Find the constant a in the function and then find the distribution function.
(ii) Find probability of $\{-1 \le X \le 1\}$ and median of X.

b) In a hamster breeder's experience, the number X of live pups in a litter of a female not over twelve months in age who has not borne a litter in the past six weeks has the probability distribution: (10+10)

X	3	4	5	6	7	8	9
$f(x)$	0.04	0.10	0.26	0.31	0.22	0.05	0.02

(i) Find the probability that the next litter will produce five to seven live pups.

(ii) Find the probability that the next litter will produce at least six live pups.

Q3. a) An urn A contains 3 white and 2 black balls, and urn B contains 5 white and 3 black balls. If an urn is chosen at random and a ball is drawn from it and found to be white, find the probability that it came from urn A.

b) In a game a player either loses $1 with a probability 0.25, wins $1 with a probability 0.4, or wins $4 with a probability 0.35. What are the expectation and the standard deviation of the winnings?

(10+10)

Q4. a) There is a probability of 0.93 that a visitor to a website will bounce (leave the website without clicking on any links). What is the probability that at least 10 of the next 12 visitors to the website will bounce? What is the expected number and variance of visitor to a website will bounce?

(b) A small lake contains 50 fish. One day a fisherman catches 10 of these fish and tags them so that they can be recognized if they are caught again. The tagged fish are released back into the lake. The next day the fisherman goes out and catches 8 fish, which are kept in the fishing boat until they are all released at the end of the day.
(i) Recognize the distribution of the number of tagged fish caught by the fisherman on the second day?
(ii) Find the probability that 3 tagged fish are caught on the second day.

(10+10)

Q5. (a) Suppose that $X \sim U(-3, 8)$. Find: (i) $E(X)$ (ii) The standard deviation of X.

(iii) The upper quartile of the distribution   (iv) $P(0 \le X \le 4)$

(b) The resistance in milliohms of 1 meter of copper cable at a certain temperature is normally distributed with mean $\mu = 23.8$ and variance $\sigma^2 = 1.28$.
(i) What is the probability that a 1-meter segment of copper cable has a resistance less than 23.0?
(ii) What is the probability that a 1-meter segment of copper cable has a resistance greater than 24.0?
(iii) What is the probability that a 1-meter segment of copper cable has a resistance between 24.2 and 24.5?

(10+10)

Handwritten annotations:

otal
—0 bish

8
—
0

$\frac{10}{50}$

$(q-p)$

$2 = \frac{\mu - \sigma}{x}$
=

0.89

ite
ack ⟩ 7 5

5 white ⟩ 8
3 black ⟩

$= {}^3C_8 \cdot {}^1C_3 + {}^1C_p \cdot {}^0C_5$

$3C_5 + {}^1C_5 + {}^5C_8 +$

$= {}^5C_3 \cdot {}^8C_2 + {}^8C_5 \cdot {}^8C_1$

${}^5C_8$

$0.8 + 56 + 8 \cdot {}^nC_r$

$= {}^5C_3 + {}^5C_1 + {}^8C_5 + {}^8C_1$